Generation of Crystal Structures of Acetic Acid and Its Halogenated Analogs

R. S. PAYNE,[1] R. J. ROBERTS,[1] R. C. ROWE,[1] and R. DOCHERTY[2]

[1] *Zeneca Pharmaceuticals, Hurdsfield Industrial Estate, Macclesfield, Cheshire SK10 2NA, UK*
[2] *Zeneca Specialties, Hexagon House, Blackley, Manchester, Cheshire SK10 2NA, UK*

Received 20 February 1997; accepted 17 June 1997

ABSTRACT: The approach of Karfunkel and Gdanitz has been used to predict possible crystal structures of acetic acid and three of its monohalogenated analogs starting with the molecular structure alone. The results demonstrate that this approach is capable of finding many, if not all, of the possible packing arrangements of molecules of this size, but that it is not currently capable of correctly ranking these structures in terms of their enthalpy. This deficiency is probably due to inadequacies in the force field used to minimize the structures. The inadequacies relate to the description of acidic hydrogen bonds and halogen–halogen interactions. © 1998 John Wiley & Sons, Inc. *J Comput Chem* **19**: 1–20, 1998

Keywords: crystal structure prediction; polymorphism; atom–atom potential method; force fields; crystal structure solution

Introduction

The crystal structure of acetic acid has been experimentally determined on a number of occasions.[1–3] Although different conditions of temperature and pressure have been studied, only small variations in the dimensions of the unit cell have been observed, the basic crystal structure remaining the same. This structure is based on two hydrogen-bonded chains of molecules related by an ''n,'' glide approximately aligned with the (0, 1, 1) and (0, 1, −1) directions. Figure 1a contains a four-molecule segment of such a chain, cut from the crystal structure of acetic acid, available as Ref. Code ACETAC10 in the Cambridge Structural Database (CSD).[4] Figure 1b contains a view down one of the chains within the crystal lattice. Formic acid has a similar chain-based structure, but virtually all other monocarboxylic acids crystallize as dimer-based structures. Figure 2 shows two views of the dimer arrangement of fluoroacetic acid (CSD Ref. Code FACETC10). The hydrogen-bonding in this family of compounds has been extensively studied, and the reason for the preference for dimers is thought to be steric hindrance in all molecules other than formic and acetic acids.[5] Sev-

Correspondence to: R. S. Payne; e-mail: robin.payne@alderley.zeneca.com

CCC 0192-8651 / 98 / 010001-20

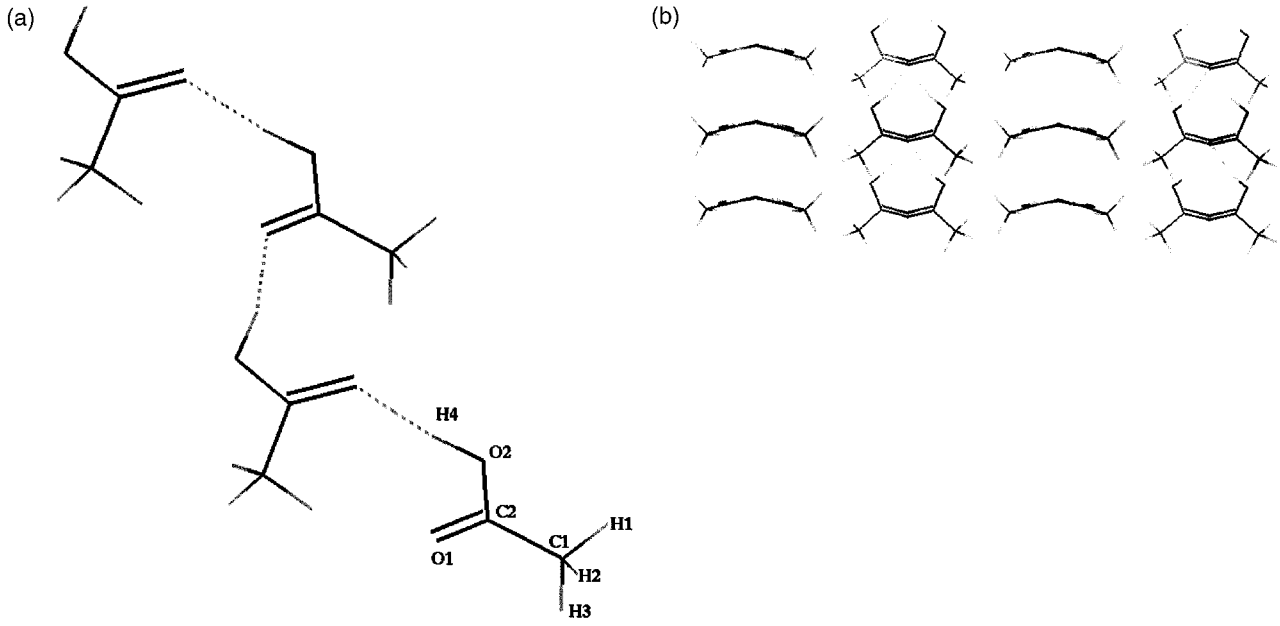

FIGURE 1. (a) The catemer motif, and (b) a view down the (0, 1, 1) direction in the experimentally known crystal structure of acetic acid.

eral investigators have demonstrated that the enthalpies of formation of the chain and dimer hydrogen bonds are very similar, with chains being, if anything, slightly the more stable arrangement.[6,7] Clearly, because the kinetic effects of crystallization are unlikely to obviate the formation of chains in the large number of experiments that have been performed on these molecules, interactions involving the atoms attached to the carboxylic acid unit must favor dimers energetically. The structure is so sensitive to these noncarboxylic acid atom–atom interactions that the mere substitution of a hydrogen atom with a halogen can swing the energy balance in favor of dimers. Support for this thesis is provided by the fact that fluoro-, chloro-, and bromoacetic acids exist as dimers in the solid state.[8–10] There is evidence of polymorphism for chloroacetic acid, where the most stable form (α) is unusually found to have a hydrogen-bonded tetramer as its basic component.[11] The metastable β-form is based on dimers. Both structures are in the space group, $P2_1/c$, α has $Z = 8$ and β has $Z = 4$. Bromoacetic acid exhibits two polymorphic forms, but both are constructed from dimers. Form I is in $P2_1/c$, with $Z = 4$ and form II is in Pccn, with $Z = 4$[10]. The three dimer structures of the monohalogenated acetic acids in $P2_1/c$ were obtained from the CSD. Views of their packing arrangements are contained in Figures 2 (flouroacetic, CSD Ref. Code FACETC10), 3 (chloroacetic, CSD Ref. Code CLACET02), and 4 (bromoacetic, CSD Ref. Code BRMACA). Despite the fundamental centrosymmetric dimer motif, the relationship between adjacent dimers differs significantly between the three molecular arrays. It is thought that these differences might arise from the combined effects of increasing Van der Waals radii, and an increasing propensity for halogen–halgen interactions with halogen atomic weight.[12]

In recent years, a number of independent attempts have been made to design molecular modeling approaches that generate or predict the crys-

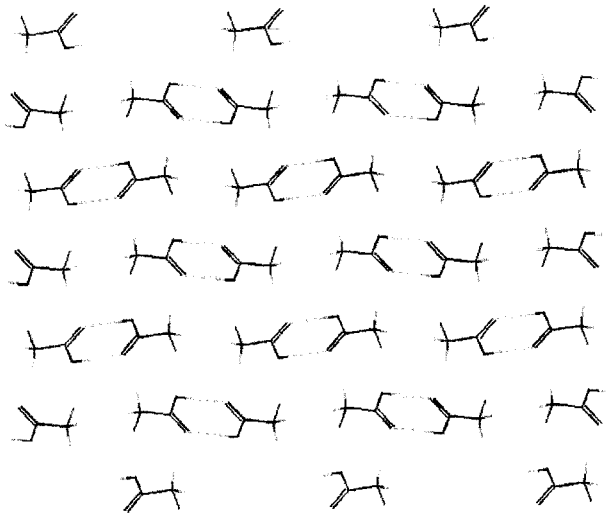

FIGURE 2. Dimers in the experimental, $P2_1/c$ crystal structure of fluoroacetic acid.

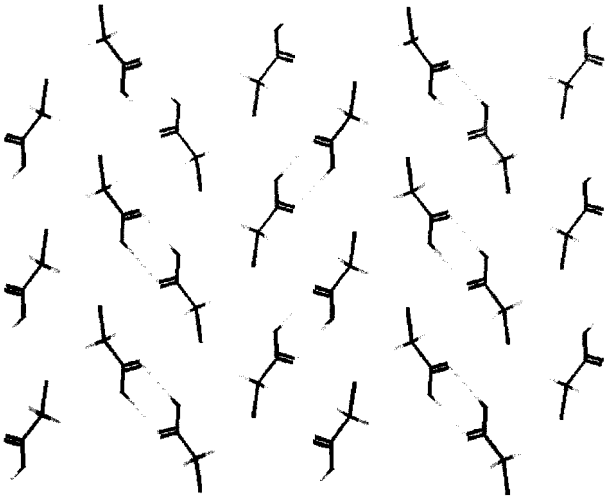

FIGURE 3. Dimers in the experimental, $P2_1/c$ crystal structure of chloroacetic acid.

tal structures of organic molecules from their molecular structures.[13–21] The motivation for this work is derived from a need to understand the phenomenon of polymorphism, in which one molecular structure can exist in more than one pure crystalline form. Different polymorphs have different physical properties, and there are several fields of endeavor that might benefit from the ability to derive conditions under which certain structures might form; for example, pharmaceuticals, pigments, electrophotographic materials and explosives. Ideally, a ''polymorph predictor'' would generate all of the possible crystal forms of a molecule, rank them in terms of free energy, and offer a measure for which structures were most likely to form under various conditions of crystallization. To date, it could be argued that we are some way from achieving this goal.[22,23]

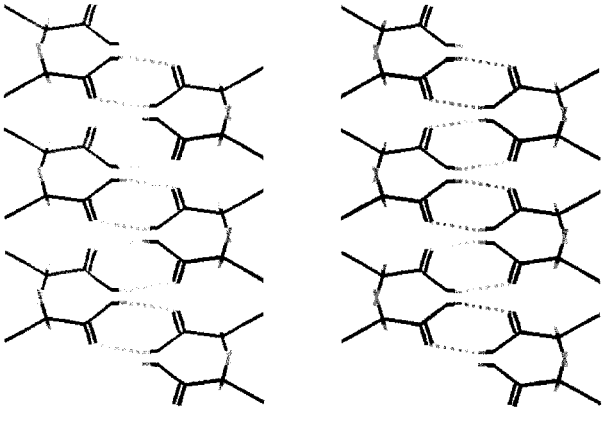

FIGURE 4. Dimers in the experimental, $P2_1/c$ crystal structure of bromoacetic acid.

One route to generating chemically reasonable crystal structures, given only the molecular structure, is by analogy. For example, Derissen and Smit[7] generated hypothetical acetic acid structures based on the dimer structure of fluoroacetic acid. They created many possible structural models by rotating the dimer arrangement around three perpendicular axes passing through the center of symmetry. The lattice energies of the resulting models were minimized using the atom–atom potential approach[24] and a special hydrogen-bond potential of the Lippincott and Schroeder form.[25] It was found that the dimer structure with the lowest lattice energy was less than 1 kcal mol^{-1} less stable than the experimentally known form. Within the errors associated with the method by which energies were calculated, where intramolecular interactions and any entropic contribution were not considered, these structures were described as equally stable.

A second approach to structure generation is to use the crystallographic knowledge that only close-packed structures are found to exist experimentally, and that only certain symmetry operations will lead to close-packed structures. A toolbox of symmetry operations are used to generate ''crystal nuclei,'' which are assembled into full crystal structures. Again, these are generally minimized using the atom–atom potential method. Gavezzotti and coworkers have been major exponents of this approach, making maximum use of their crystallographic knowledge and their own force-field parameters.[16,26] However, in the case of monocarboxylic acids, they would probably generate a nucleus based on a dimer before assembling crystals. Thus, the *a priori* observation that the vast majority of monocarboxylic acids form dimers in the solid state would probably not lead to a successful prediction of the crystal structures of formic and acetic acids.[13]

The third, and potentially most general approach to the problem is to generate all of the possible packing arrangements in all of the available space groups. This is, of course, relatively costly in computational terms, and efficient programming and expediency must be employed if results are to be gained in a reasonable time frame. For example, the number of space groups investigated might be limited to those which account for the vast majority of the known crystal structures of organic molecules.[12] C^2 Polymorph[27] is a commercially available program based on the approach first described by Gdanitz[17] and exploited by Kar-

funkel, Gdanitz, and coworkers.[19,28,29] Possible packing arrangements are generated by a Monte Carlo/simulated annealing (MC/SA) search of free space. Similar arrangements are clustered together and represented by the lowest energy structure in the cluster. The clustering process reduces the number of structures, which are subsequently minimized using the atom–atom potential method. Clustering by similarity is repeated after minimization, and the resulting structures are presented in order of their total energy (that total energy including intra- and intermolecular contributions). This approach is a major step forward in the field of structure prediction, but there are still a number of deficiencies which leave the prediction process some distance from the aforementioned ideal. The most expensive part of the process, in terms of cpu time, is that of minimizing the potential crystal structures. Chaka et al. have recently proposed a way of reducing the time required at this stage, by using the molecular mutipole moments to locate potential packing arrangements, prior to minimization with an atom–atom potential.[30]

In this article, we use acetic acid and its halogenated analogs to test the performance of C^2 Polymorph. There are a number of references to the polymorphism of acetic acid,[31–34] but only one such study could be described as containing firm evidence that this molecule might exist in more than one crystal form.[35] However, the fact that relatively small changes in molecular structure lead to significant changes in the crystal structure for this family of molecules provides a serious test for the performance of any program designed to generate or predict crystal structures. This application demonstrates the great potential of the approach, while exemplifying some of its current deficiencies.

Experimental

PREDICTIVE METHOD

This section will constitute a summary of detailed descriptions of the process published elsewhere.[19]

The generation of feasible crystal structures for a particular molecule is a demonstration of the ability to solve the problem of locating the global energy minimum, along with all of the low energy local minima. Generally, polymorphs that are experimentally realizable have structures within ca. 1 kcal mol^{-1} of the global minimum. Thus, the location of all possible low energy structures would be termed ''the prediction of all polymorphs,'' even though it is not currently possible to simulate real crystallization processes by which crystal nuclei form and grow.

The initial step in the approach of Karfunkel and Gdanitz is to employ MC/SA[36] to locate crude possible packing arrangements. An oversized unit cell is constructed and molecules are placed into it according to the symmetry demands of a particular space group. The molecules are then moved randomly, using one of the 11 angular variables defined for this process.[19]

Subsequently, structures with bad contacts between the molecules are rejected and energy minimization is performed on the accepted structures, keeping all variables constant except the unit cell parameters and the distance between the centers of the molecules. Restricting the MC moves to angular variables prevents the ''evaporation'' of the model crystals, which is a known problem with preceding MC/SA approaches to the packing problem. The SA temperature is increased quite rapidly, until a certain number of sequential structures, say 12, are accepted via the Metropolis criterion.[37] At this stage, the system is said to have attained a temperature which makes all moves possible, and hence that a ''melt'' has been achieved. Cooling then takes place at a relatively low rate, and the Metropolis acceptance criterion is applied following each MC move. Cooling ceases when a prescribed number of structures are rejected, indicating that no moves are possible, and that the system is ''frozen.'' There are three important parameters that can be set by the user at the MC/SA stage of the method:

- N_accept—The number of structures which are accepted for the system to be considered a random ''melt.'' High values of N_accept ensure that all energy barriers are surmountable, but also lead to a greater number of structures for subsequent minimization.
- Heat_Factor—Defines the rate of heating during the attainment of a ''melt.''
- Cool_Factor—Similarly defines the rate of cooling. This parameter is particularly important, because slow cooling allows more MC moves per unit temperature, and a higher probability of finding all of the low energy structures. However, low cooling rates again generate more structures for minimization.

The values used for these parameters, and others governing the predictive method, are listed in Table I.

The MC/SA procedure can generate many thousands of crude structures, a significant number of which will be similar. Similar structures are identified according to the Cerius[2] program manual, by ''examining the partial radial distribution functions between pairs of force field atom types of the two structures being compared.''[27] This step is said to be an acceptable means of reducing the number of structures that require full structural minimization, as long as the structures are sufficiently similar to fall into the same local energy minimum. The important parameter for clustering is termed ''Tolerance.'' The lower the Tolerance, the more similar a structure must be to the reference structure to be deemed a member of the same cluster. It is advised that a moderate setting can be used after the MC/SA search, but that the finest settings must be used after minimization, when quite similar structures may represent different polymorphs (see Table I).

Full structural minimization is performed relaxing all of the degrees of freedom of the crystal system, including the conformational flexibility of the molecules and the unit cell parameters. Minimization terminates after a user-defined level of ''RMSF,'' the root mean squared force for convergence (see Table I).

The minimization procedure utilizes the atom–atom potential method, and the force field ''Dreiding 2.21'' was chosen for this work. Dreiding 2.21 is a generic force field, which takes explicit account of hydrogen bonding.[38] Karfunkel and Gdanitz have demonstrated that the results of their approach are best if high level quantum mechanical (QM) calculations are performed to derive atomic point charges which fit the electrostatic potential of the chosen molecular conformation (ESP charges). In this case, charges were calculated for the crystal conformation of each molecule, using a 6-31G** basis set, and the Merz–Singh–Kollman (MSK) method for deriving the charges. The QM calculations were performed using Gaussian-92.[39] Other, perhaps more sophisticated, schemes for deriving ESP charges were available in Gaussian-92, but the MSK scheme was the best with appropriate radii for atoms up to bromine in the periodic table.[40]

TABLE I.
Values Used for Important User-Definable Parameters in All Structure Prediction Runs Described in this Study.

Parameter	Value
N_accept	12
Heat_Factor	0.025
Cool_Factor	0.002
Tolerance	0.15
RMSF	0.001

It has been recommended that the molecular structures are optimized to high levels using *ab initio* calculations, before charges are calculated.[19] This recommendation was the subject of a test in which crystal structures were predicted using three different starting points: (a) the molecular structure of acetic acid from the crystal structure solution; (b) that structure after low level QM optimization; and (c) that structure after high level QM optimization. This test is described further in subsequent sections.

ACETIC ACID

Optimization and Crystal Structure Prediction in $Pna2_1$

Three predictive runs on acetic acid were performed in the space group $Pna2_1$. These were identical apart from slight differences in the molecular conformations. Run A used the crystal conformation, run B used a STO-3G-optimized version of this structure, and run C a 6-31G**-optimized version. 6-31G** MSK charges were generated for each of these three starting conformations. The aim of these runs was to reveal any significant effects which might result from optimizing the molecule before structure prediction. Crystal structures with each of the optimized conformations were rebuilt in the experimentally known unit cell, using a best-fit procedure to locate the center of gravity of the new conformers as close to the position of the original molecule as possible. These structures were minimized to investigate the effect of optimization on the performance of the minimization process. The magnitude of the difference between the experimental and minimized unit cell parameters is frequently used as a means of evaluating the adequacy of the force field used in the atom–atom potential method.

Structure Prediction in Other Space Groups

The molecular conformer from the crystal structure solution was also used in predictive runs in

the space groups $P2_1/c$, $C2/c$, $P\bar{1}$, $P2_1$, and $P2_12_12_1$. If we include $Pna2_1$, these six space groups account for approximately 80% of the structures in the CSD.[12,41] This set of predictions was made to investigate the packing arrangements likely in each case, and to derive the relative energies of these arrangements. These space groups would be the preferred ones for an investigation of a molecule for which the crystal structure was unknown.

HALOGENATED ANALOGS OF ACETIC ACID

The crystal structure conformers of fluoro-, chloro-, and bromoacetic acid, with 6-31G** ESP MSK charges, were used to find possible structures in space groups $Pna2_1$ and $P2_1/c$. In the case of fluoroacetic acid, an additional run was performed in $P2_1/c$, using the 6-31G**-optimized version of the crystal molecular structure, to further investigate the effect of molecular optimization. The basic premise was to investigate the relative stability of the chain and dimer structures generated in each case, with the knowledge that, experimentally, chains of acetic acid are observed in $Pna2_1$ and dimers of the halogenated molecules are found in $P2_1/c$.

Results and Discussion

Optimization of the molecular structure of acetic acid and fluoroacetic acid resulted in very minor changes to the geometry of the structure. Subsequent calculation of ESP charges also showed minor differences resulting from optimization. Figure 1a features a labeled molecule indicating the atoms referred to in Tables II and III. These tables contain some of the geometric parameters of the optimized acetic acid structures, and corresponding charges, respectively. Table IV contains the unit cell param-

TABLE II.
Geometric Parameters for Experimental Molecular Structure, and the Two Optimized Structures of Acetic Acid.[a]

Structure	O1—C2—O2	C1—H1	C1—H2	C1—H3	C1—C2	C2=O1	C2—O2	O2—H4
Crystal structure	122.0	1.115	1.082	1.066	1.482	1.219	1.319	1.002
Minimized crystal structure	116.5	1.0869	1.0858	1.0851	1.5100	1.1900	1.3908	0.9763
STO-3G	121.8	1.0849	1.0864	1.0864	1.5368	1.2168	1.3917	0.9900
6-31G**	122.4	1.0795	1.0840	1.0840	1.5011	1.1872	1.3309	0.9480

[a] The atom labels refer to Figure 1a. The angle is in degrees and bond lengths are in angstroms.

TABLE III.
MSK ESP Charges Generated for Molecular Structures of Acetic Acid Used in This Study.[a]

Structure	C1	C2	H1	H2	H3	H4	O1	O2
Crystal structure	0.8784	−0.4440	0.4563	0.1152	0.1484	0.1478	−0.6653	−0.6368
STO-3G	0.8264	−0.4623	0.4364	0.1322	0.1579	0.1579	−0.6558	−0.5927
6-31G**	0.8864	−0.4813	0.4508	0.1331	0.1527	0.1527	−0.6725	−0.6219

[a] The atom labels refer to Figure 1a.

TABLE IV.
Unit Cell Parameters and Energies of Acetic Acid Structures Minimized After Molecular Optimization to Different Levels.[a]

Structure	*a*	*b*	*c*	ρ	Total energy
Crystal structure	13.31	4.09	5.77	1.270	—
Crystal structure minimized	14.32	4.00	5.34	1.303	−123.55
STO-3G minimized	14.20	4.00	5.42	1.295	−119.75
6-31G** minimized	14.41	3.99	5.33	1.302	−125.98

[a] Unit cell vector lengths are given in angstroms, density in grams per cubic centimeter, energy in kilocalories per mole. The space group is orthorhombic, $Pna2_1$.

eters of the known crystal structure of acetic acid, and the values for each of the minimized structures.

Within Cerius,[2] is is possible to apply the clustering algorithm described previously, to give a measure of the similarity between a predicted structure and a given crystal structure. Structures with low similarity values (i.e., ≤ 0.1) can be considered virtually identical. One might therefore define a successful structure prediction to be one in which a structure was generated for which the similarity measure was less than this value. Ideally, we would have this low level measure of similarity on comparison with the CS, but given the fact that all predicted structures are the result of force-field minimization, we choose to regard a successful prediction as one in which a structure has a similarity measure of ≤ 0.1 when compared with the CS_min. Thus, as demonstrated by column 3 of Table Va–c, all three runs with different

TABLE V.
Results of Predicting Structures of Acetic Acid in Space Group $Pna2_1$.[a]

Frame number	Sim. vs. CS	Sim. vs. CS_min	*a*	*b*	*c*	ρ	Total energy	PA
(a) Run A: crystal structure conformer								
1	0.320	0.327	5.58	8.90	6.07	1.322	−124.47	1
2	0.307	0.337	9.40	6.32	5.23	1.285	−123.74	1
3	0.253	0.255	5.75	9.26	6.02	1.244	−123.73	1
4	0.229	0.229	9.91	6.93	4.66	1.246	−123.67	3
5	0.219	0.214	7.68	8.33	4.87	1.279	−123.54	3
6[b]	0.197	0.013	14.32	4.00	5.34	1.303	−123.54	2
7	0.204	0.199	13.56	4.26	5.39	1.281	−123.53	2
8[c]	0.194	0.201	13.45	4.29	5.38	1.284	−123.53	2
9	0.226	0.216	9.90	7.07	4.58	1.244	−123.49	3
10	0.310	0.332	6.73	7.58	6.43	1.215	−123.48	1
(b) Run B: STO-3G conformer								
1	0.312	0.323	5.58	8.89	6.09	1.319	−120.59	1
2	0.322	0.331	9.40	6.33	6.24	1.280	−119.92	1
3	0.214	0.220	9.69	7.14	4.65	1.241	−119.89	3
4[c]	0.199	0.147	14.30	4.00	5.39	1.293	−119.88	2
5	0.258	0.267	5.60	9.42	6.10	1.241	−119.85	1
6	0.201	0.203	13.38	4.30	5.43	1.276	−119.83	2
7	0.219	0.217	7.62	8.31	4.93	1.277	−119.82	3
8[b]	0.206	0.000	14.20	4.00	5.42	1.295	−119.76	2
9	0.325	0.336	6.84	7.48	6.42	1.215	−119.73	1
10	0.223	0.235	9.91	5.39	6.03	1.238	−119.72	3
(c) Run C: 6-31G** conformer								
1	0.314	0.324	5.54	8.95	6.08	1.322	−126.78	1
2	0.323	0.332	9.38	6.33	5.23	1.284	−126.10	1
3[b,c]	0.198	0.091	14.46	3.99	5.31	1.300	−126.06	2
4	0.214	0.218	13.70	5.12	4.53	1.257	−126.03	3
5	0.224	0.231	9.87	6.97	4.66	1.243	−125.92	3
6	0.209	0.193	13.56	4.26	5.39	1.280	−125.90	2
7	0.222	0.221	7.66	8.34	4.89	1.277	−125.87	3
8	0.210	0.192	5.44	14.32	4.01	1.278	−124.86	3
9	0.210	0.212	10.19	5.35	5.89	1.243	−125.82	3
10	0.219	0.209	5.12	14.06	4.48	1.236	−125.82	3

[a] Unit cell vector lengths are given in angstroms, densities (ρ) in grams per cubic centimeter, and total energies in kilocalories per mole. CS refers to the known crystal structure and CS_min to the minimized crystal structure. Frame numbers are allocated to predicted structures ranked in terms of total energy. PA refers to the packing arrangements described in the text and Table VI.
[b] Structures most similar to the minimized known crystal structure for each conformer in each run.
[c] Structures most similar to the known crystal structure in each run.

levels of molecular optimization could be termed successful. Each located a structure with a similarity measure ≤ 0.1 when compared with the CS_min. In Table V, it can be seen that all of the acetic acid runs performed in the experimentally known space group, $Pna2_1$, ranked the known structure as less stable than several others. Table V contains a number of other sets of data: the length of the unit cell vectors and the density of the structures and their total energies. Total energy equates to sum of all the atom–atom interaction energies in a structure: it constitutes a calculation of the enthalpy, summing that associated with the molecular structure and that associated with the intermolecular structure.

Although there are subtle differences between the structures predicted in each of the three runs, there are strong similarities in the results. These are clear in the lengths of the unit cell vectors and, particularly, in the packing arrangements (PA). Here, the PA will be primarily described in terms of a graph set, which classifies the molecular topology of hydrogen-bonding motifs in a molecular array.[42–44] Additional descriptive terms will also be used. These are aimed at differentiating between arrays which would have an identical graph set, but which have different symmetry elements between the molecules. While these elements are properties of the space groups for the various predictive runs, the additional terms for each PA are descriptive, allowing groups of similar structures to be better defined. These groups are described in Table VI, and illustrated in the figures listed (Fig. 5). Table VI contains descriptions that summarize structures observed for acetic acid and additional entries that will be used in reviewing the results for the halogenated acetic acids.

Predictive runs for the acetic acid crystal molecular structure were performed in five other space groups, namely, $P2_1/c$, C2/c, P-1, $P2_1$, and $P2_12_12_1$. The energies of the 20 most stable structures located in these predictions are compared with those in $Pna2_1$ in Figure 6. It is clear that $P2_1/c$ yields the highest number of low energy structures, although some of these are virtually identical. Despite this, the lowest energy structure was located in each of the three space groups $P2_1/c$, $P2_12_12_1$, and $Pna2_1$. All structures in these space groups were reranked in terms of their density, and the 20 most dense in each plotted in Figure 7. The densest structures were generated in $P2_12_12_1$ and $Pna2_1$. However, after the third entry in each data set, $P2_1/c$ gave the densest structures, as might be expected given that this space group is the one most commonly adopted by organic molecules.

Each of the halogenated structures were packed in space groups $Pna2_1$ and $P2_1/c$. For fluoroacetic acid an extra run was performed using an optimized molecular structure in $P2_1/c$. The ten lowest energy structures for each of the runs on the halogenated structures are presented in Table VII. In cases where the structure most similar to the known crystal structure was not one of the ten most stable structures, it is included in an extra row of the table. Entries are not made where data are irrelevant or calculations were not possible.

It is encouraging that the structure generation process of C^2 Polymorph found structures that were similar to the minimized known crystal structures of the molecules in every experimental run. Structures with the packing arrangements of the known crystal structures were found to exist within the ten most stable structures in each case. To confirm the value of the similarity measure, frame 6 of run A on acetic acid was subjected to Rietveld refinement against the simulated powder pattern of the known crystal structure.[45] The program used for refinement is known as DBWS.[46] One limitation of this package is that chemical constraints (e.g., the retention of certain bond lengths and angles) cannot be applied to the refinement process. This limitation meant that it was not possible to undertake extensive refinement and yet retain sensible molecular geometries, and should be noted when judging the extent to which refinement proved to be possible. Initially, the unit cell parameters were refined alone, but subsequently the coordinates of the heavy atoms were included in the refinement process. After approximately 500 iterations, the R-factor for frame 6 fell from 98% to 22.5%. Although it is debatable whether this R-factor indicates a structure solution of acceptable quality, it demonstrates that, given the limitations of simulated powder data, frame 6 could have been used to give a structural solution for the known form of acetic acid. A comparison of the powder patterns of the refined structure and the crystal structure is given in Figure 8.

Within the ten lowest energy structures of runs A, B, and C (Table Va–c), there are structures with PA2 (equivalent to the CS) that exhibit low similarity measures versus the CS and the CS_min, but these only coincide for run C (Table Vc, frame 3). This observation is supportive of the recommendation that high level QM optimization is advantageous when using the method of Karfunkel

and Gdanitz to predict crystal structures. However, the degree of success achieved in the runs which used less optimized geometries suggests that the benefits of expending large amounts of cpu time optimizing a molecule are minor, at least for an example as simple as acetic acid.

The differences in energy between the structures most similar to the crystal structure and the lowest energy structures generated in runs A, B, and C are 0.93, 0.83, and 0.67 kcal mol^{-1}, respectively. These results indicate that the known CS is approximately 0.8 kcal mol^{-1} less stable than a structure with PA1. It is expected that any experimentally observed structure will be the global minimum (lowest energy structure) or exist within about 1 kcal mol^{-1} of that global minimum. It will

TABLE VI.
Packing Arrangements (PA) Discussed in This Article.

PA	Figure	Graph set	Additional terms	Description
1	5a	C(4)	Parallel x2	Chains of molecules with all heavy atoms coplanar. Adjacent molecular planes are offset by half a molecule.
2	1b	C(4)	Herring x2	Two sets of identical chains which bent at the axis of the hydrogen bonds. The known form of acetic acid has this arrangement.
3	5b	C(4)	Helical x2	Two sets of helical chains, with adjacent chains staggered. Arrangement looks like dimers when viewed down the axis of the chains.
4	3	$R_2^2(8)$	x2	Two sets of dimers inclined such that a view down the axis of one set leaves a view of the other set as dimers. Compare with known crystal structure of chloroacetic acid.
5	5c	C(4)	Helical x1	One set of helical chains. $P2_1/c$
6	5d	C(4)	Buckled out x2	Similar to arrangement 1, but with the chains bent at the axis of the hydrogen bonds. Degree of buckle increases with the size of the halogen. The buckle results in the C halogen bond pointing outward.
7	5e	C(4)C(2)	Ladder	Two motifs creating ladder chains.
8	4	$R_2^2(8)$	x1	Similar to PA4 except that viewing the arrangement down the axis of one set of dimers results in an end view of the other set. Compare with the known dimer structures of fluoro- and bromoacetic acids.
9	5f	C(4)	Parallel x1	Planar chains with no half-molecule offset between planes.
10	5g	C(2)		Chains formed by only the O—H ··· O—OH bond seen in PA7. The C(4) motif exists, but at a greater distance than the 2.5-Å limit set for describing hydrogen bonds in this work.
11	5h	C(4)	Buckled in x2	Similar to PA6, but with the C-halogen bonds pointing inward.
12	5i	C(4)	Buckled anti-x2	Similar to PA6, but with the C-halogen bonds pointing in opposite directions in sequential molecules down the chains.

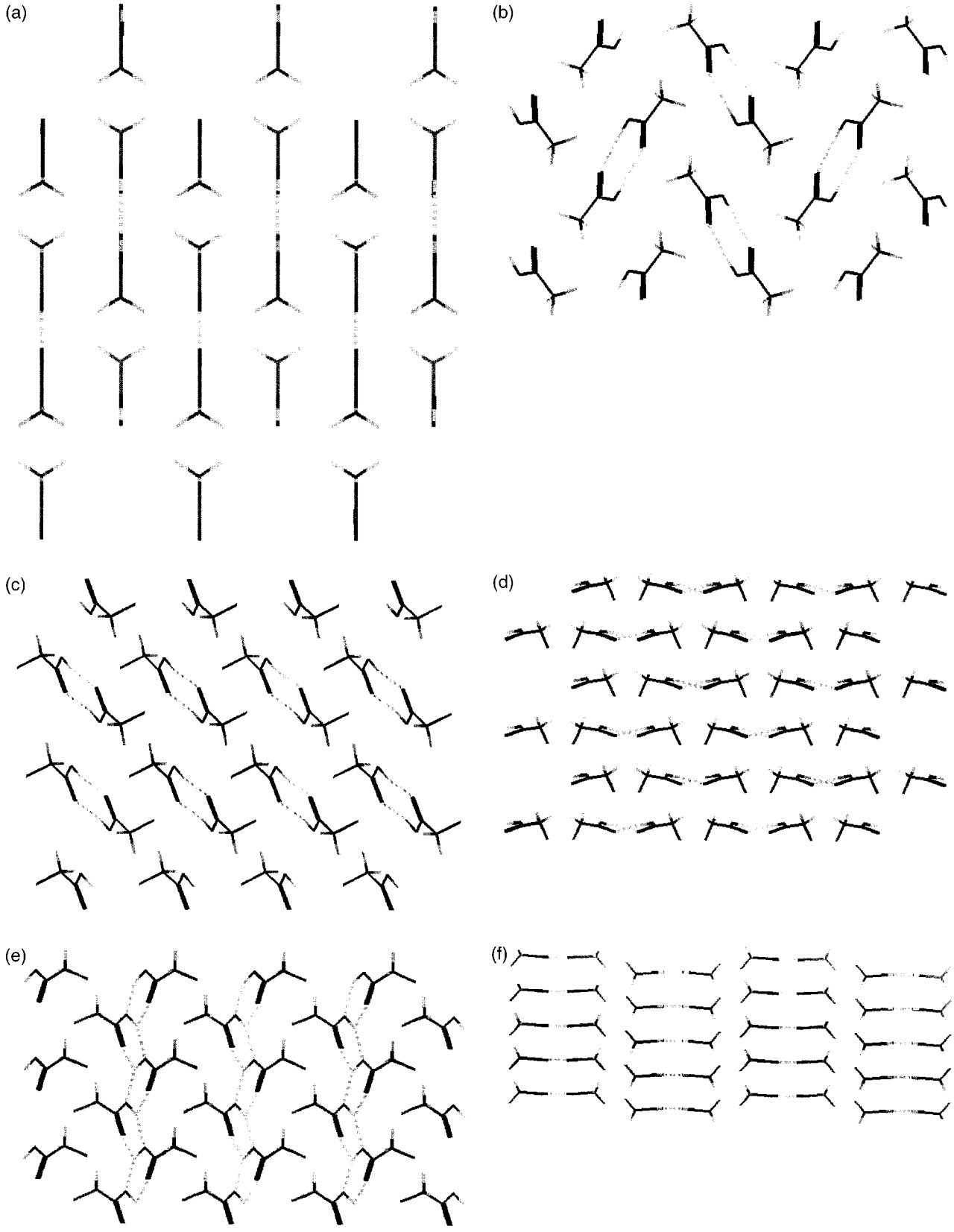

FIGURE 5. Structural views exemplifying some of the packing arrangements listed in Table VI.

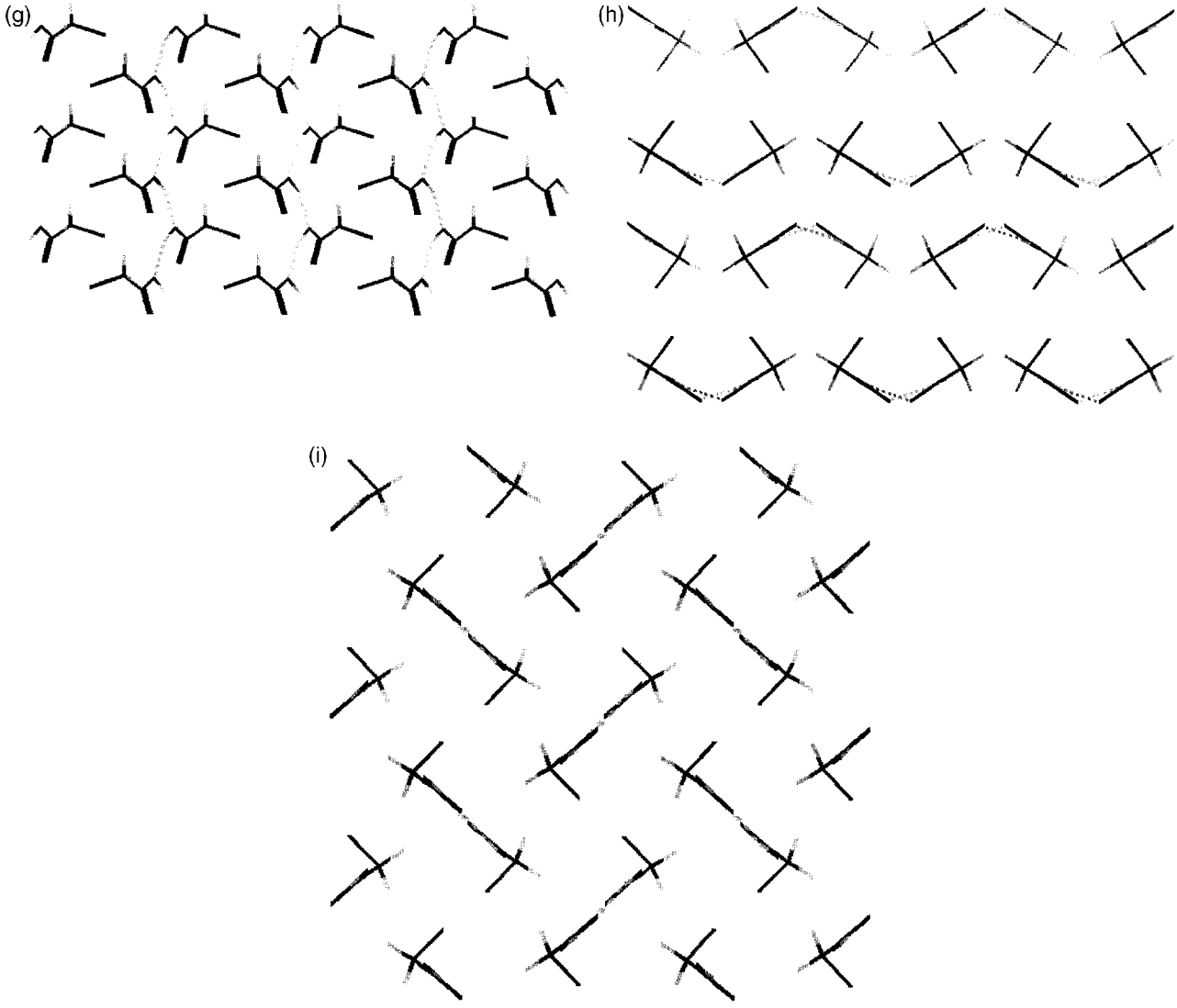

FIGURE 5. continued

be shown later than we must be wary of claiming great accuracy for a generic force field used to calculate these energy differences. However, consideration of Table V suggests that PAs 1, 2, and 3 ought to be experimentally observable.

Is it possible that one of the predicted structures agreed with the powder data of the low temperature, high pressure form of acetic acid reported by Bertie and Wilton[35]? Answering this question did not prove easy, partly because the original powder data was not available to us, and partly because we believed the unit cell derived from this data to be incorrect. In an attempt to solve these issues, the *d*-spacings from Bertie and Wilton's study were indexed using the program TREOR90.[47] The solution was as follows: $a = 5.67$ Å, $b = 13.19$ Å, $c = 3.90$ Å, $\beta = 93.63°$, and cell volume 291.09 Å^3 and shall henceforth be referred to as form II of acetic acid. The figure of merit for this result was $M = 12$, with one line unindexed. The unindexed line was one of the weaker ones in the data at a *d*-spacing of 2.655 Å. Two of the unit cell vectors are approximately half as long as those derived by Bertie and Wilton, who suggested that there were $Z = 16$ molecules in the unit cell. In our solution, a molecular volume of ca. 73 Å^3 would indicate $Z = 4$, a much more likely situation for a monoclinic space group. Consideration of the indexed lines indicated systematic absences consistent with the space group $P2_1/c$. The unit cell dimensions of all of the acetic acid structures predicted in $P2_1/c$ were reviewed. Frame 21 had values most similar to those

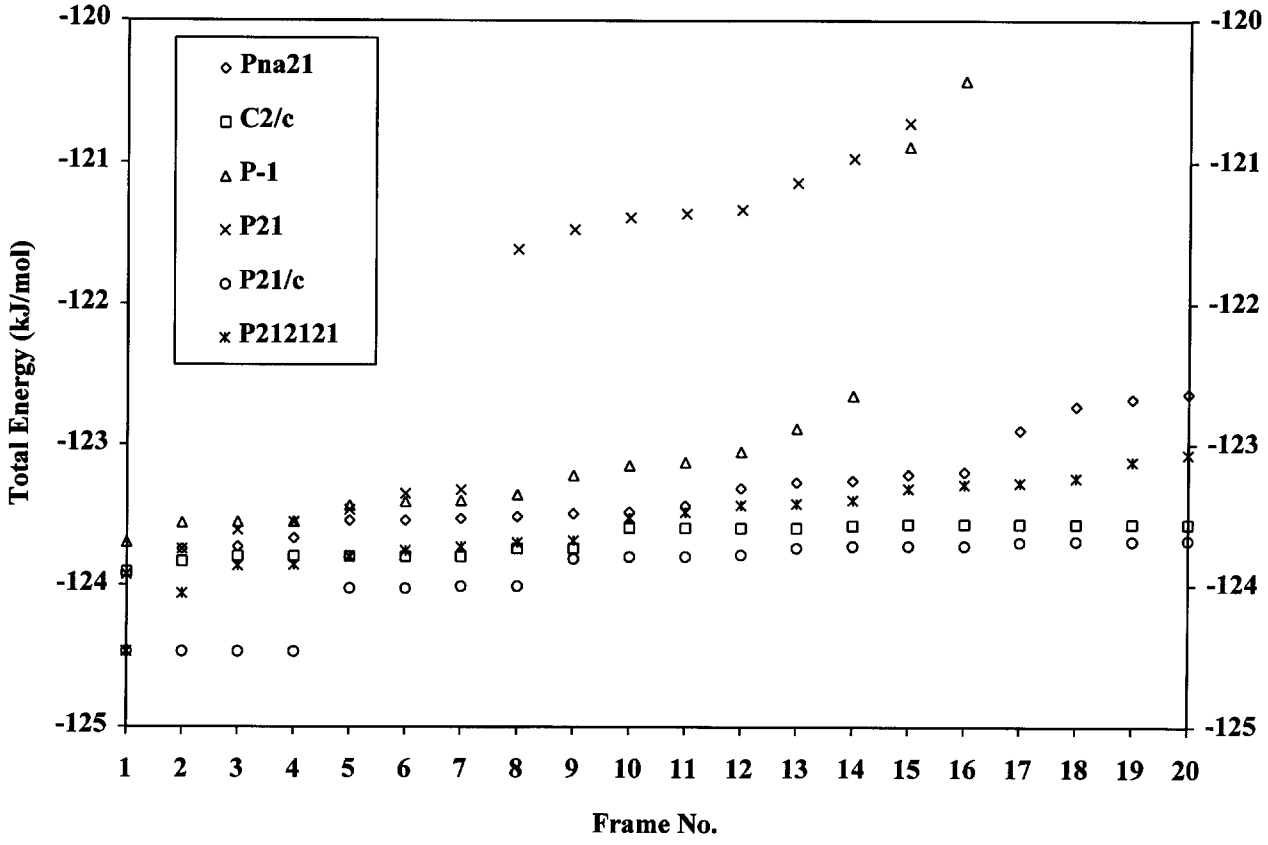

FIGURE 6. Energies of the 20 most stable structures located in each of the six space groups chosen for structure predictions.

quoted above, $a = 5.35$ Å, $b = 13.79$ Å, $c = 4.20$ Å, $\beta = 94.83°$. Rietveld refinement was applied to frame 21, using the simulated powder pattern of Bertie and Wilton's data as the experimental data to refine against. The results of about 500 refinement cycles are shown in Figure 9. Given the nature of the data and the fact that atomic positions could not sensibly be used in the refinement process, the similarity between the two patterns suggests that frame 21 is a good representation of the crystal structure.

At first sight the dimensions of the form II unit cell are similar to those of form I, and the generation of a monoclinic cell could be attributed to the pressure applied to the system when this structure was generated. However, frame 21 indicates that the PA for form II is different. The graph set is the same, C(4), but all of the chains run parallel to each other: there are not two sets as seen in form I. Figure 10 contains a diagram of the structure of frame 21, which should be compared with Figure 1b. The generation of a structure that is similar to Bertie and Wilton's form II is a powerful demonstration of a potential secondary use of the structure prediction program: the solution of crystal structures from powder data. This aspect of science has been pursued by a number of workers, because there are often problems in obtaining sufficiently large crystals of important compounds for single crystal x-ray diffraction to provide a structural solution. However, the technical difficulty of peak overlap has meant that there has been little success in the field of organic molecules.[48] Computational methods are likely to bridge the gap between the molecular structure and the experimental powder data, for an increasing number of organic compounds in the near future.[49–51]

A review of the predictive run of the CS conformer of acetic acid in space group $P2_1/c$ was particularly interesting in that dimer structures were generated. The lowest energy structure generated in this run was identical to one of those low energy structures generated in $Pna2_1$, with PA1. The lowest energy dimer structure was only ca. 0.4 kcal mol^{-1} less stable than this apparent global minimum. This result agrees with previous find-

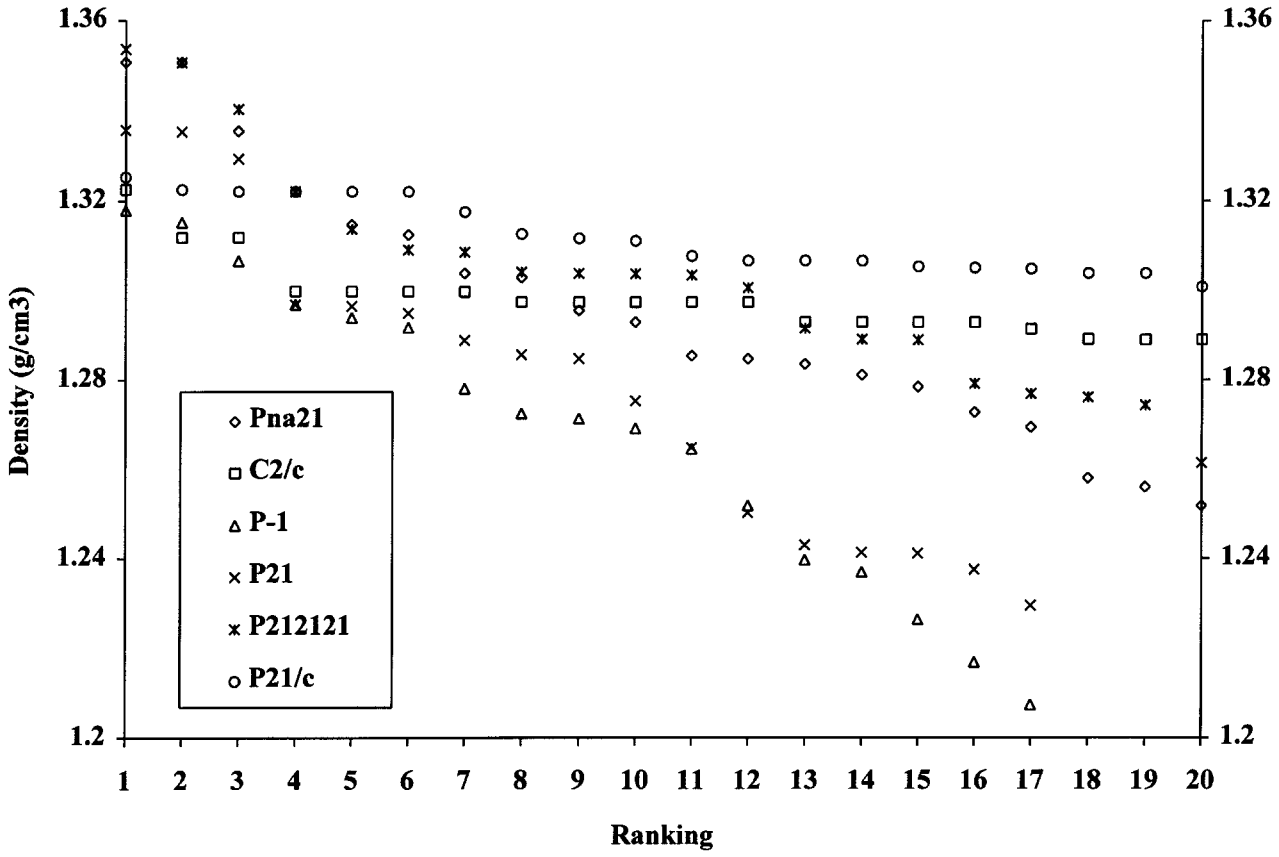

FIGURE 7. Densities of the 20 most stable structures located in each of the six space groups chosen for structure predictions.

ings that dimer structures are energetically similar to chains for acetic acid, and implies that dimers should also be experimentally observable.[6,7] No packing arrangements other than numbers 1, 2, 3, and 4 were seen in predictions for this molecule.

In $P2_1/c$, structure prediction for fluoroacetic acid was as might have been expected, given that the experimental CS contains dimers: a dimer-containing structure was found to be the most stable PA (the results of runs on the halogenated acetic acids are summarized in Table VII). Despite this, there were anomalies. The structure most similar to the CS, occurring within the ten most stable predictions, was ranked energetically less stable than structures with PA1. Furthermore, there was another structure substantially more like the CS that was 1.1 kcal mol^{-1} higher in energy than the predicted global minimum (frame 27). This situation was not significantly improved, qualitatively or quantitatively, by using the 6-31G**-optimized conformer in the prediction. In this case, a structure similar to that of the CS was found eighth on the list, but the most similar structure of all was frame 26, again ca. 1.1 kcal mol^{-1} less stable than the predicted lowest energy structure. In $Pna2_1$, PA1 provided the most stable form. For acetic acid, this arrangement had been the most stable overall; for fluoroacetic acid, it was the second most stable overall, but lay very close to the lowest energy dimer form. Two new structural variations were seen in this prediction: PA6 and PA7. PA6 appeared to represent a distortion to PA1, resulting from accomodation of the fluorine atom. PA7 was the only structure observed in this work that contained more than one hydrogen bond motif, having two types, which combined to produce a ladder-like arrangement. There were no structures which contained fluorine–fluorine distances shorter than twice the Van der Waals radius of this atom.

When chloroacetic acid was packed in $P2_1/c$, it was surprising to find that the three most stable structures were based on chains: apparently similar to those seen for acetic and fluoroacetic acids, but with additional distortions induced by the presence of the chlorine atom. The lowest energy

TABLE VII.
The Ten Most Stable Structures for Each of the Predictive Runs Performed on Halogenated Acetic Acid Structures.[a]

Frame number	Sim. vs. CS	Sim. vs. CS_min	*a*	*b*	*c*	β	ρ	Total energy	PA
(a) Fluoroacetic acid: Crystal structure conformer in $P2_1/c$									
CS	—	—	4.31	7.61	10.03	94.74	1.581	—	8
CS_min	—	—	4.46	7.42	10.48	94.62	1.500	−37.50	8
1	0.258	0.246	8.50	7.96	4.98	73.63	1.605	−38.60	8
2	0.326	0.325	6.63	7.72	6.22	89.98	1.627	−38.58	1
3	0.339	0.345	6.22	7.72	6.63	89.98	1.627	−38.58	1
4	0.271	0.249	4.98	7.96	8.50	106.32	1.604	−38.57	8
5[b,c]	0.240	0.210	9.99	8.09	4.37	109.83	1.558	−38.26	4
6	0.275	0.248	4.31	16.86	4.86	70.43	1.559	−38.26	4
7	0.276	0.258	4.86	16.87	4.31	109.54	1.559	−38.24	4
8	0.258	0.246	10.15	5.98	5.40	95.09	1.586	−38.18	5
9	0.241	0.228	7.58	5.86	8.26	68.21	1.524	−38.15	4
10	0.302	0.286	4.05	20.58	4.12	97.57	1.523	−38.04	9
27[d]	0.212	0.063	4.46	7.43	11.05	70.79	1.500	−37.52	9
(b) Fluoroacetic acid: 6-31G** conformer in $P2_1/c$									
CS_min	—	—	4.47	7.45	10.47	94.80	1.493	−45.35	8
1	0.263	0.249	8.51	7.96	4.98	73.73	1.599	−46.43	8
2	0.323	0.328	6.62	7.74	6.24	90.00	1.622	−46.43	1
3	0.342	0.338	6.62	7.74	6.24	90.00	1.622	−46.43	1
4	0.338	0.351	6.24	7.74	6.62	90.00	1.622	−46.43	1
5	0.278	0.263	9.32	7.02	5.34	107.50	1.556	−46.38	8
6	0.271	0.255	4.88	16.80	4.32	109.92	1.558	−46.13	4
7	0.272	0.248	4.88	16.80	4.32	70.08	1.558	−46.13	4
8[b,c]	0.233	0.219	10.00	8.09	4.39	69.97	1.556	−46.10	4
9	0.354	0.336	10.00	8.09	4.39	69.97	1.555	−46.10	4
10	0.292	0.288	4.04	20.51	4.13	97.06	1.526	−46.01	9
25[d]	0.209	0.005	4.47	7.45	11.12	117.12	1.493	−45.35	9
(c) Fluoroacetic acid: crystal structure conformer in $Pna2_1$									
1	—	—	6.63	7.72	6.22		1.627	−38.58	1
2	—	—	5.62	9.06	6.61		1.541	−37.96	6
3	—	—	10.64	7.28	4.48		1.492	−37.61	3
4	—	—	7.78	6.56	6.56		1.550	−37.49	1
5	—	—	10.31	6.18	5.26		1.546	−37.47	1
6	—	—	7.95	10.22	4.60		1.388	−37.42	3
7	—	—	8.53	7.43	5.31		1.541	−37.39	3
8	—	—	5.375	10.17	6.16		1.539	−37.25	1
9	—	—	12.08	5.56	4.90		1.576	−37.14	7
10	—	—	10.42	7.748	4.27		1.507	−37.06	3
(d) Chloroacetic acid: crystal structure conformer in $P2_1/c$									
CS	—	—	4.64	12.97	6.69	107.93	1.641	—	4
CS_min	—	—	4.61	11.26	7.87	112.81	1.666	−116.06	4
1	0.250	0.254	6.62	7.03	8.57	109.66	1.672	−117.29	11
2	0.266	0.238	5.54	5.53	13.33	106.88	1.606	−117.05	5
3	0.292	0.283	10.66	4.82	7.96	110.61	1.639	−116.90	9
4[c]	0.255	0.228	7.76	4.69	10.99	104.10	1.619	−116.89	4
5[b]	0.247	0.232	4.66	14.33	5.92	74.51	1.647	−116.86	4
6	0.255	0.228	7.89	5.68	8.96	71.02	1.655	−116.84	4
7	0.299	0.294	8.12	4.19	11.69	102.00	1.613	−116.62	5

TABLE VII.
(continued)

Frame number	Sim. vs. CS	Sim. vs. CS_min	*a*	*b*	*c*	β	ρ	Total energy	PA
8	0.252	0.257	4.34	11.49	7.84	93.17	1.607	−116.54	4
9	0.302	0.322	4.34	11.49	7.84	86.83	1.607	−116.54	4
10	0.267	0.263	4.34	11.49	7.84	86.82	1.607	−116.54	4
29[d]	0.242	0.007	7.42	11.25	4.61	77.88	1.666	−116.06	4
(e) Chloroacetic acid: crystal structure conformer in $Pna2_1$									
1	—	—	10.17	7.96	4.82		1.609	−117.05	3
2	—	—	12.93	5.80	5.28		1.584	−116.75	3
3	—	—	20.41	4.44	4.54		1.528	−116.14	3
4	—	—	10.73	7.63	4.999		1.537	−116.00	3
5	—	—	13.34	5.76	5.01		1.632	−115.86	10
6	—	—	9.61	6.61	6.35		1.555	−115.81	12
7	—	—	6.40	7.15	8.409		1.635	−115.80	11
8	—	—	13.92	6.349	4.57		1.555	−115.76	3
9	—	—	6.59	7.71	7.69		1.608	−115.74	6
10	—	—	8.22	7.05	6.81		1.591	−115.73	3
(f) Bromoacetic acid: crystal structure conformer in $P2_1/c$									
CS	—	—	12.64	4.34	8.01	109.88	2.234	—	8
CS_min	—	—	12.82	4.33	7.75	110.61	2.290	−82.28	8
1	0.324	0.312	6.74	7.16	8.70	70.58	2.333	−83.88	11
2	0.332	0.324	12.88	5.83	5.44	93.77	2.267	−83.63	11
3	0.330	0.315	7.71	6.83	7.64	98.94	2.323	−83.45	5
4	0.344	0.333	5.66	7.07	10.34	81.09	2.257	−83.31	12
5	0.330	0.315	8.87	11.85	4.27	115.74	2.284	−83.29	4
6	0.329	0.313	6.64	14.05	4.80	116.95	2.314	−83.25	4
7[b,c]	0.235	0.200	8.31	4.16	11.95	79.62	2.269	−83.22	5
8	0.346	0.338	5.33	10.37	7.65	111.78	2.333	−83.16	4
9	0.342	0.339	7.65	10.37	5.37	111.77	2.333	−83.16	4
10	0.328	0.315	9.22	4.25	10.55	96.24	2.248	−83.07	5
69[d]	0.197	0.081	13.43	4.33	7.75	63.32	2.289	−82.28	8
(g) Bromoacetic acid: crystal structure conformer in $Pna2_1$									
1	—	—	10.23	8.04	4.91		2.284	−83.50	3
2	—	—	12.94	5.81	5.55		2.210	−83.11	3
3	—	—	12.95	5.86	5.50		2.213	−83.09	3
4	—	—	13.14	7.00	4.57		2.195	−82.95	3
5	—	—	6.60	7.29	8.32		2.306	−82.84	11
6	—	—	14.21	6.47	4.52		2.220	−82.84	3
7	—	—	13.38	4.37	6.96		2.266	−82.82	3
8	—	—	9.77	6.65	6.49		2.189	−82.68	12
9	—	—	6.97	7.63	7.66		2.263	−82.65	6
10	—	—	10.57	8.84	4.30		2.293	−82.51	3

[a] Additional entries are made where the structure most similar to the known crystal structure is not one of the ten most stable predicted. Unit cell vector lengths (a, b, c) are given in angstroms, densities (ρ) in grams per cubic centimeter, and total energies in kilocalories per mole per cell. CS refers to the known crystal structure and CS_min to the minimized crystal structure. Frame numbers are allocated to predicted structures ranked in terms of total energy.

[b] The structure most similar to the CS in the ten structures predicted to be most stable.

[c] The structure most similar to the CS_min in the ten structures predicted to be most stable.

[d] The structure most similar to the CS and CS_min, if ranked outside the ten most stable predictions.

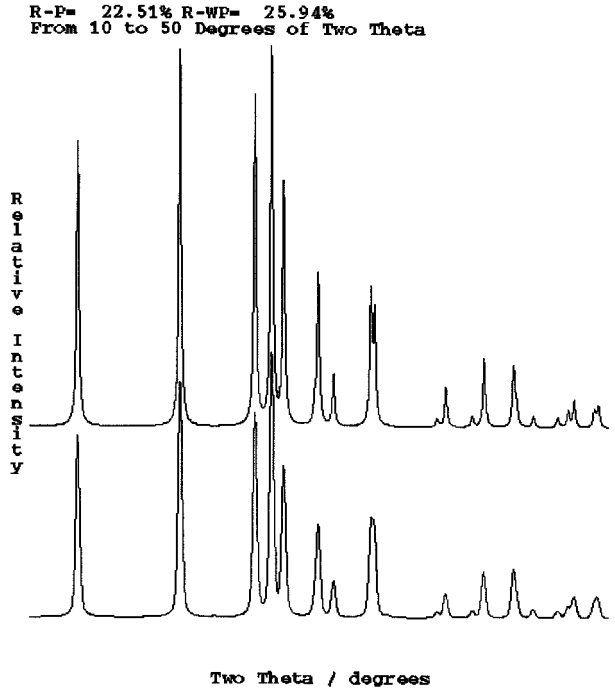

FIGURE 8. A comparison of a simulated powder pattern for the known crystal structure of acetic acid (upper pattern), with the pattern for frame 6 of acetic acid run A, after Rietveld refinement (lower pattern).

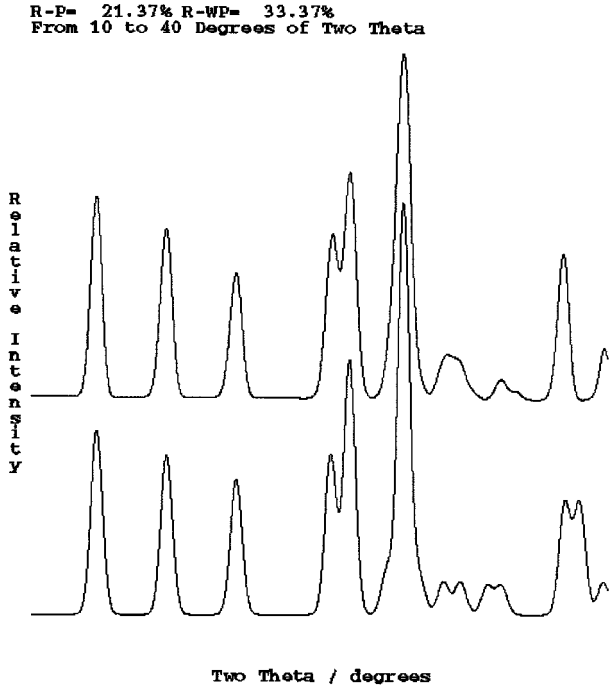

FIGURE 9. Rietveld refined powder data for frame 21 of acetic acid run in space group $P2_1/c$ (upper pattern), compared with the simulated powder pattern generated by convoluting a Gaussian peak shape with the *d*-spacings reported by Bertie and Wilton, for a low temperature, high pressure structure[35] (lower pattern).

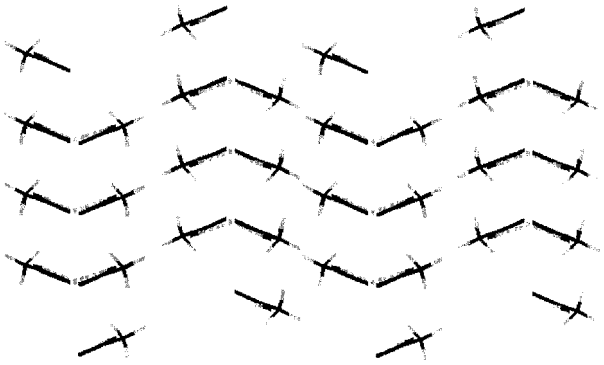

FIGURE 10. The crystal structure of frame 21 of acetic acid predicted in space group $P2_1/c$. This structure gives a good representation of the experimental data reported by Bertie and Wilton.[35]

dimerized structure, (frame 4), some 0.4 kcal mol^{-1} above the predicted global minimum, was similar to the CS_min, and was followed by another frame similar to the CS. However, as with fluoroacetic acid, the structure most similar to the CS was ranked well down the list at frame 29, some 1.2 kcal mol^{-1} from the global minimum. The prediction in $Pna2_1$ produced a number of low energy structures with the helical arrangement of PA3. There was no evidence of PA1, probably because the chlorine atom sterically favors helices over parallel chains. PA10 was observed: this arrangement was similar to the ladder structure seen in the analogous prediction for fluoroactic acid, but the hydrogen bond distance for the C(4) motif was 2.7 Å, which was greater than the 2.5 Å used to define such bonds in this study. Two other buckled chain arrangements were see: PA11, where the halogen atoms pointed into the concave shape seen when looking down the chain; and PA12, with one halogen pointing into and one pointing out of the concavity. There was evidence of chlorine–chlorine distances shorter than twice the Van der Walls radius of this atom.

The results for bromoacetic acid were qualitatively similar to those obtained for chloroacetic acid. Prediction in $P2_1/c$ found dimer structures to be about 0.4 kcal mol^{-1} less stable than chain structures. In this case, however, the lowest energy dimer structures were found not to be very similar to the CS. Indeed, the structure most similar in the entire prediction was ranked at frame 76, with similarity measures against the CS and the CS_min that were well in excess of the value 0.1, surpassed in all other predictions. The predictive process was

decidedly less effective for bromoacetic acid than for the other molecules described in this work.

It is difficult where the major source of error lies in the predictive process. From identical repeated runs on the acetic acid molecule (data not presented here), it is clear that the level of MC/SA structural search used in this article reproducibly located the same set of structures for this size of molecule, despite the random nature of moves made within it. In addition, the MC/SA search succeeded in finding structures that were very similar to the known CS in every run. These observations suggest that the scheme for locating possible structures is adequate, but that either the force-field parameterization, or the value of the ESP-derived charges, are at fault in terms of deriving the relative energies of the located structures.

Generating evidence that tests the accuracy of the ESP charge calculations is not easy. Such charges are sensitive to molecular conformation. In addition, charge differences lead to small differences in the values of bond lengths and angles upon force-field minimization, which can have a significant effect on conformational energy, depending on the scale of the force constants in the force field. In general, it is best to calculate charges using the formation adopted at the time of force-field parameterization. In the case of Dreiding 2.21, no charges were specified at this time and it is accepted that 6-31G** ESP charges represent the most accurate means of dealing with the electrostatic interactions when using this force-field.

One test of the force field is to minimize the CS and compare the lattice parameters of the CS and the CS_min. Differences of less than 5% are said to represent a good result. Comparison of these parameters in this work indicated that the Dreiding 2.21 force field was acceptable, according to this criterion. However, for the molecular systems investigated here, the strength of hydrogen bonds and the weaker halogen-related interactions may be critical if the relative energies of structures are to be calculated to an accuracy of better than 0.5 kcal mol^{-1}, as is required to predict the stability ranking of polymorphs. Drieding 2.21 indicated that the energy associated with each hydrogen bond in the acetic acid catemer structure was -3.74 kcal mol^{-1}, whereas the hydrogen bonds in the dimer structures of fluoro-, chloro-, and bromoacetic acid were -3.72, -3.67, and -3.20 kcal mol^{-1}, respectively. Hydrogen bonds in a dimer created from two molecules of acetic acid had an energy of -3.71 kcal mol^{-1} associated with them. These calculations are of a magnitude that is consistent with the results of other workers,[6,7] and suggest that small changes in the lengths of these bonds would be sufficient to significantly change the order of stability of structures in these molecular systems. Such changes might well occur within the $\pm 5\%$ change in unit cell parameters considered acceptable when minimizing a CS. In their work on force-field paramaterization for hydrogen bonds, Hagler et al. reported that the hydrogen bond length in acetic acid increased from 1.64 to 1.69 Å upon application of their force field.[6] Dreiding 2.21 led to a change from 1.65 Å (as measured from the CS during this work) to 1.91 Å. This is clearly a large increase, placing the distance in the minimized structure at the very upper limit of the distribution of hydrogen bond lengths found for carboxylic acids in the CSD (see Fig. 11). This force field does not provide a good description of the hydrogen bond in molecules containing the carboxylic acid group. Table VIII contains the hydrogen bond and halogen–halogen distances for each of the structures reported in this work, and indicates the effect of minimization on them. It is evident that there is no advantage in applying high level *ab initio* optimization to these structures. The hydrogen bond lengths in acetic acid, fluoroacetic acid, and bromoacetic acid are far greater in the CS_min's than in the CSs. For fluoro- and chloroacetic acids, the halogen–halogen distances are also significantly increased by minimization.

Although the hydrogen bonds of carboxylic acids are the strongest interactions between molecules in the solid state, they nearly always result in dimer formation. As a consequence, types of interaction that might otherwise be termed secondary (.e.g, halogen–halogen interactions), can dominate the crystal packing arrangement.[12] This is evident in bromoacetic acid, where dimers are arranged in planes with interplanar bromine–bromine interactions involving distances of less than twice the Van der Waals radius of bromine. In chloroacetic acid, the effect is not quite so clear, but there is one chlorine–chlorine distance less than twice the Van der Waals radius. Fluorine-substituted molecules do not, in general, show fluorine–fluorine interactions, in fact there is some evidence the fluorine atoms may repel each other in some systems.[12] Even so, the mere fact of their high electronegativity means that they will have a significant role to play in the packing of carboxylic acid dimers. These observations suggest that the halogen–halogen distances are critical to energy calculations in these systems.

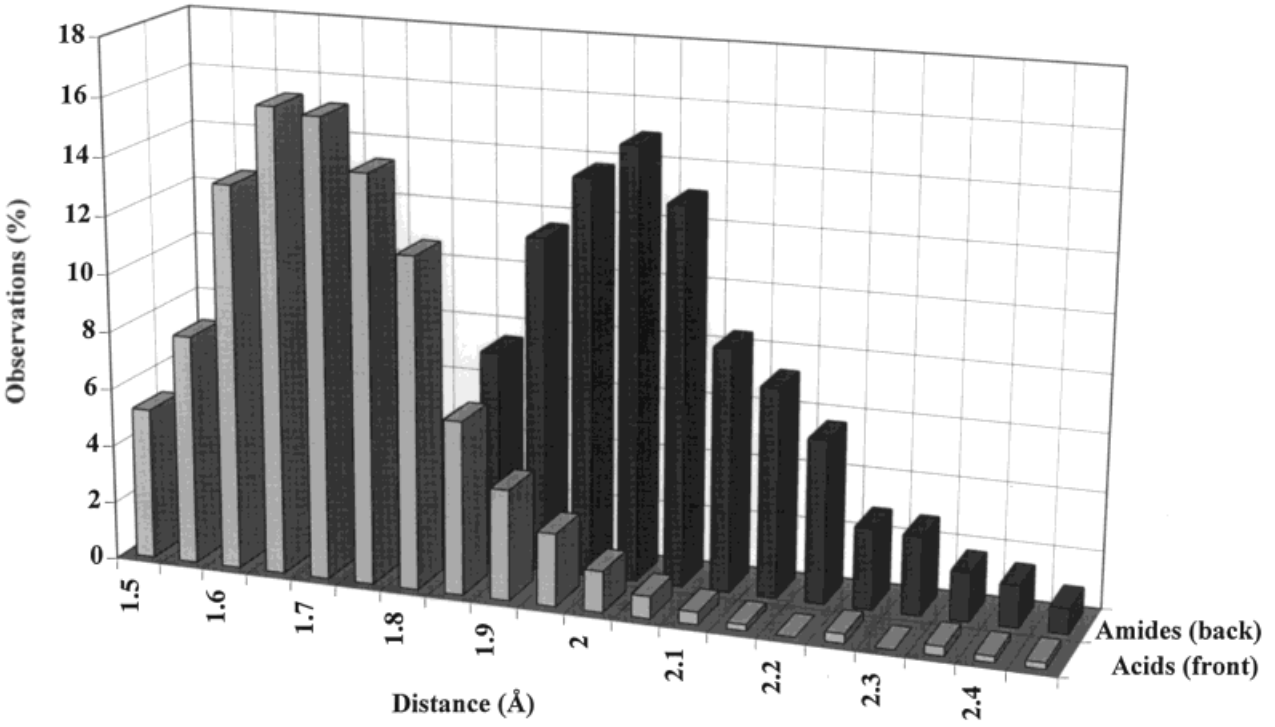

FIGURE 11. The observed distribution of hydrogen bond distances for carboxylic acid dimers located in the CSD. The search considered organic molecules (no metal-containing structures were considered), for which the structure solution gave an *R*-factor of ≤ 0.07. The mean bond lengths were 1.71 Å for carboxylic acids and 2.10 Å for amides.

In summary, we must conclude that Dreiding 2.21 does not represent an ideal way of using ESP charges within the atom–atom potential method. The main reason for this probably arises from the fact that Dreiding 2.21 was parameterized on molecular structures that had been extracted from their crystal structures, and so cannot be said to account for intermolecular interactions. In addition, the bond distances were derived from Van der Waals atomic radii, and we know that halogen–halogen contacts (along with other ''special'' interatomic interactions) are seldom consistent

TABLE VIII.
Hydrogen Bond and Halogen–Halogen Distances for the CSs and Force-Field-Minimized CSs (CS_min's) Considered in This Article.[a]

Acid molecule	Conformer	Structure	H bond distance	Shortest halogen–halogen distance
Acetic	CS	CS	1.651	—
Acetic	CS	CS_min	1.905	—
Acetic	STO-3G	CS_min	1.917	—
Acetic	6-31G**	CS_min	1.909	—
Fluoroacetic	CS	CS	1.733	3.174
Fluoroacetic	CS	CS_min	1.911	3.230
Fluoroacetic	6-31G**	CS_min	1.916	3.265
Chloroacetic	CS	CS	1.918	4.014
Chloroacetic	CS	CS_min	1.917	4.552
Bromoacetic	CS	CS	1.896	3.717
Bromoacetic	CS	CS_min	1.985	3.708

[a] Distances are measured from hydrogen to oxygen in the case of hydrogen bonds. All values are given in angstroms.

with the chronicled values. The parameterization of hydrogen bonds was based on molecular interactions in water, and, as such, unlikely to represent a general parameterization for all organic molcules. In fact, it has been shown the Dreiding 2.21 provides a preferred hydrogen bond length consistent with amine-containing structures, rather than carboxylic acids. This is shown in Figure 12, where the general potential curve for hydrogen bonding is shown for Dreiding 2.21, compared to specific amide and acidic potential functions derived for intermolecular interactions.[52,53] The minimum for Dreiding 2.21 is close to the mean for amides, but is too long for acids. One way of improving this situation would be to set a different preferred hydrogen bond length for carboxylic acid structures, or, the definition of more than one type of hydrogen bond for molecular structures capable of forming more than one type. However, the key to successful crystal structure, and hence polymorph prediction, is the construction of a unified force field that is as good at describing intermolecular interactions as it is at describing molecular structures.

Despite its limitations, Drieding 2.21 is, as its authors state, generic: it can be applied to a wide range of heteroatomic molecules with reasonable accuracy, but to none with high accuracy.[38] It is one of the few force fields that explicitly accounts for hydrogen bonds in some way, and for which the crystal field can be treated with the efficient method of Ewald summation.

In this article, we have focused on the prediction of structures for a set of small molecules with corresponding limited conformational flexibility. Even so, limitations of the force field used for minimization have become evident. Further limitations are likely to be manifested when the technique is applied to larger molecules with increased flexibility. One can envisage a need to investigate the packing of a number of conformers if *ab initio* work is to offer a realistic assessment of polymorphism for such systems. The problems of conformational polymorphism remain the subject of considerable debate and recent work has illustrated the complexity of considering this phenomenon theoretically.[54].

Each predictive run performed on the eight-atom molecules in this study took approximately 1 day to complete on a SG Indigo^2 200-MHz machine. Similar calculations on 30- and 72-atom molecules took approximately 3 and 7 days, respectively. Although these periods of time are not unmanageable, and computational speeds are increasing, it is still necessary to invest considerable cpu time in running these calculations for medium and large molecules.

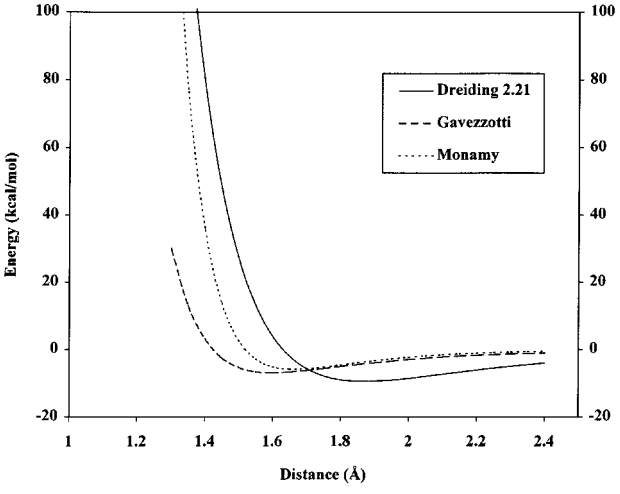

FIGURE 12. The hydrogen bonding potential of Dreiding 2.21 compared to the hydrogen bonding potentials of intermolecular force fields of Monamy et al.[52] and Gavezzotti and Fillippini.[53]

Conclusions

The methodology developed by Karfunkel and Gdanitz, and implemented in C2 Polymorph, represents a significant step forward in the prediction of crystal structures for organic molecules. For molecules of the size of acetic acid, the MC/SA search routine is highly effective in locating potential packing arrangements. For acetic acid itself, the method has located an alternative form that appears refinable against x-ray powder diffraction data reported elsewhere. This observation demonstrates a powerful alternative function for the method; that is, as an aid to solving crystal structures of molecules for which large single crystals are difficult to obtain.

Unfortunately, the state of the art in generic force fields that must be employed to minimize these potential arrangements is not sufficiently sophisticated to account for different types of hydrogen bond, or halogen–halogen interactions. As a consequence, it is not currently possible to have confidence in the relative stability of the predicted structures.

Acknowledgments

The authors thank J. Starbuck for his contribution to searching the CSD, and S. Maginn (Daresbury Laboratory) and F. Leusen (MSI) for their useful comments.

References

1. R. E. Jones and D. H. Templeton, *Acta Cryst*., **11**, 484 (1958).
2. I. Nahringbauer, *Acta Chem. Scand*., **24**, 453 (1970).
3. P. G. Johnsson, *Acta Cryst*., **B27**, 893 (1971).
4. Cambridge Crystallographic Data Centre, *QUEST3D*, Union Road, Cambridge CB21 EZ, UK, 1995.
5. L. Leiserowitz, *Acta Cryst*., **B32**, 775 (1976).
6. A. T. Hagler, P. Dauber, and S. Lifston, *J. Am. Chem. Soc*., **101**, 5131 (1979).
7. J. L. Derissen and P. H. Smit, *Acta Cryst*., **A33**, 230 (1977).
8. G. Roelofsen, J. A. Kanters, and P. Brants, *Cryst. Struct. Commun*., **7**, 313 (1978).
9. J. A. Kanters, G. Roelofsen, and T. Feenstra, *Acta Cryst*., **B32**, 3331 (1976).
10. L. Leiserowitz and D. vor der Bruck, *Cryst. Struct. Comm*., **4**, 647 (1975).
11. J. A. Kanters and G. Roelofsen, *Acta Cryst*., **B32**, 3328 (1976).
12. G. R. Desiraju, In *Crystal Engineering: The Design of Organic Solids*, Elsevier, Amsterdam, 1989.
13. B. P. van Eijck, W. T. M. Mooij, and J. Kroon, *Acta Cryst*., **B51**, 99 (1995).
14. R. A. Catlow, J. M. Thomas, C. M. Freeman, P. A. Wright, and R. G. Bell, *Proc. R. Soc. Lond*., **A442**, 85 (1993).
15. Y.-L. Change, M.-A. West, F. W. Fowler, and J. W. Lauher, *J. Am. Chem. Soc*., **115**, 5991 (1993).
16. A. Gavezzotti, *J. Am. Chem. Soc*., **113**, 4622 (1991).
17. R. J. Gdanitz, *Chem. Phys. Lett*., **190**, 391 (1992).
18. J. R. Holden, Z. Du, and H. L. Ammon, *J. Comput. Chem*., **14**, 422 (1993).
19. H. R. Karfunkel and R. J. Gdanitz, *J. Comput. Chem*., **13**, 1171 (1992).
20. J. Perlstein, *J. Am. Chem. Soc*., **114**, 1955 (1992).
21. T. Shoda, K. Yamahara, K. Okazaki, and D. E. Williams, *J. Mol. Struct*., **313**, 321 (1994).
22. A. Gavezzotti, *Acc. Chem. Res*., **27**, 309 (1994).
23. J. Maddox, *Nature*, **335**, 201 (1988).
24. A. J. Pertsin and A. I. Kitaigorodsky, *The Atom–Atom Potential Method*, Springer, Berlin, 1987.
25. R. Schroeder and E. R. Lippincott, *J. Am. Chem. Soc*., **61**, 921 (1957).
26. A. Gavezzotti, *Acta Cryst*., **B52**, 201 (1996).
27. Biosym/Molecular Simulations, Inc., *Cerius* 2, Release 2.0, 16 New England Executive Park, Burlington, MA 01803, 1995.
28. H. R. Karfunkel and F. J. J. Leusen, *Speedup J*., **6**, 43 (1992).
29. H. R. Karfunkel, F. J. J. Leusen, and R. J. Gdanitz, *J. Comp.-Aided Mat. Design*, **1**, 177 (1993).
30. A. M. Chaka, R. Zaniewski, W. Youngs, C. Tessier, and G. Klopman, *Acta Cryst*., **B52**, 165 (1996).
31. W. Tan, K. A. Krieger, and J. G. Miller, *J. Am. Chem. Soc*., **74**, 6181 (1952).
32. C. Rigaux, *C. R. Acad. Sci. Paris*, **238**, 783 (1954).
33. R. E. Jones and D. H. Templeton, *Acta Cryst*., **11**, 484 (1958).
34. R. J. Jakobsen, Y. Mikawa, and J. W. Brasch, *Spectrochim. Acta*, **23A**, 2199 (1967).
35. J. E. Bertie and R. W. Wilton, *J. Chem. Phys*., **75**, 1639 (1981).
36. M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods*, John Wiley & Sons, New York, 1986.
37. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys*., **21**, 1087 (1953).
38. S. L. Mayo, B. D. Olafson, and W. A. Goddard III, *J. Phys. Chem*., **94**, 8897 (1990).
39. M. J. Frisch et al., *Gaussian* 92/*DFT*, Revision G.1, Guassian Inc., Pittsburgh, PA, 1993.
40. M. A. Spackman, *J. Comput. Chem*., **17**, 1 (1996).
41. V. K. Belsky and P. M. Zorkii, *Acta Cryst*., **A33**, 1004 (1977).
42. M. C. Etter, *J. Phys. Chem*., **95**, 4601 (1991).
43. M. C. Etter and J. C. MacDonald, *Acta Cryst*., **B46**, 256 (1990).
44. J. Bernstein, R. E. Davis, L. Shimoni, and N.-L. Chang, *Angew. Chem. Int. Ed. Engl*., **34**, 1555 (1995).
45. R. A. Young, In, *The Rietveld Method*, R. A. Young, Ed., Oxford University Press, Oxford, 1993, p. 1.
46. See p. 30 of Young[45] for a review of popular Rietveld refinement programs.
47. P.-E. Werner, L. Erikkson, and M. Westdahl, *J. Appl. Cryst*., **18**, 367 (1985).
48. A. K. Cheetham, In *The Rietveld Method*, R. A. Young, Ed., Oxford University Press, Oxford, 1993, p. 276.
49. M. Tremayne, B. M. Kariuki, and K. D. M. Harris, *J. Appl. Cryst*., **29**, 211 (1996).
50. R. B. Hammond, K. J. Roberts, R. Docherty, M. Edmondson, and R. Gairns, *J. Chem. Soc. Perkin Trans*., 2, **7**, 1527 (1996).
51. H. R. Karfunkel, Z. J. Wu, A. Burkard, G. Rihs, D. Sinnreich, H. M. Buerger, and J. Stanek, *Acta Cryst*., **B52**, 555 (1996).
52. F. A. Monamy, L. M. Carruthers, R. F. Mcguire, and H. A. Scheraga, *J. Phys. Chem*., **78**, 1595 (1974).
53. A. Gavezzotti and G. Filippini, *J. Phys. Chem*., **98**, 4831 (1994).
54. D. Buttar, M. H. Charlton, R. Docherty and J. Starbuck, in press.